# eCommerce Data Science Assignment

Prepared by: Shabes Khan

Date: January 27, 2025

## Table of Contents

## Task 1: Exploratory Data Analysis (EDA)

Overview:

This task involved analyzing the dataset to uncover patterns, trends, and actionable insights.

Key Insights:

1. South America has the highest number of customers, representing a major market.

2. Electronics and Books are the top-performing categories by revenue.

3. Monthly sales trends reveal seasonal spikes, ideal for targeted campaigns.

4. Top customers account for significant revenue, suitable for loyalty programs.

5. Price-sensitive products (lower price, higher quantity sold) dominate certain segments.

Visualizations:

1. Customer distribution by region (bar chart).

2. Monthly sales trends (line chart).

3. Product popularity vs. pricing (scatter plot).

## Task 2: Lookalike Model

Overview:

This task involved building a lookalike model to identify similar customers based on profile and transaction data.

Methodology:

1. Features such as total spend, average transaction value, and preferred categories were engineered.

2. Cosine similarity was used to measure similarity between customers.

3. The top 3 lookalike customers for each target customer were identified.

Sample Result:

For Customer C0001, the top 3 similar customers were:

1. Customer C0072 (Similarity Score: 0.9988)

2. Customer C0190 (Similarity Score: 0.9976)

3. Customer C0043 (Similarity Score: 0.9965)

## Task 3: Customer Segmentation

Overview:

This task involved clustering customers into groups to identify behavioral patterns.

Clustering Approach:

1. Algorithm: KMeans

2. Optimal Number of Clusters: 4 (determined using the Davies-Bouldin Index)

3. Evaluation Metric: DB Index = 0.44 (indicating well-separated clusters).

Cluster Characteristics:

1. Cluster 0: Moderate spenders, focus on Home Decor (44 customers).

2. Cluster 1: High spenders, focus on Books (54 customers).

3. Cluster 2: Balanced spenders, focus on Electronics (51 customers).

4. Cluster 3: Moderate spenders, focus on Clothing (50 customers).

Insights:

1. Cluster 1 (High-Value Customers) should be prioritized for retention strategies.

2. Seasonal promotions can target Cluster 0 (Home Decor enthusiasts).

3. Bundle offers for Cluster 2 (Electronics enthusiasts) can drive higher value.

## Conclusion and Recommendations

This assignment demonstrates the ability to derive business insights, build predictive models, and perform customer segmentation.

Strategic Recommendations:

1. Focus on high-value customers (Cluster 1) with loyalty programs.

2. Tailor promotions for seasonal trends and regional preferences.

3. Use the lookalike model for personalized marketing and customer acquisition.

These insights and models can help the company optimize customer engagement and maximize revenue.